European
Commission

Data For Policy 2017 Conference

Big Data Test Infrastructure

European Commission, DIGIT D1

Roberto Barcellan,
Head of Unit DIGIT.D1 – Data Services

Marco Fichera,
Project Manager

London, 7 September 2017

Agenda

Big Data Test Infrastructure

Background & Objectives

Approach and status of the study

Task 1 - Use cases and requirements analysis

Task 2 - Data ontologies and data exchange APIs

Task 3 - Design of the infrastructure

Next Steps

Agenda

Big Data Test Infrastructure

Background & Objectives

Approach and status of the study

Task 1 - Use cases and requirements analysis

Task 2 - Data ontologies and data exchange APIs

Task 3 - Design of the infrastructure

Next Steps

Big Data Test Infrastructure

Background & Objectives

This Big Data Test Infrastructure is one of the activities under **ISA²
Action 2016.03 – Big Data for Public Administrations**, funded
by the ISA² Programme. The **ISA² programme** supports the
development of digital solutions that enable public administrations,
businesses and citizens in Europe to benefit from interoperable
cross-border and cross-sector public services.

Contact
Roberto Barcellan
Tanya Chetcuti
Marco Fichera
DIGIT D1, EC

Problem the BDTI tries to solve

- **Adoption of analytics technologies is lagging** behind in public administrations.
- **Lack of a turn-key cloud environment** that offers a full stack of technologies to test the value of new ways of processing big data

Objective

Support the execution of Big Data pilots under **different policy domains** by any interested Public Administration and Institution in Europe enabling the implementation of both Big Data pilots at National/Local level and at cross-border level.

European Commission

Big Data Test Infrastructure
Background & Objectives

Title	Problem	Solution	
Adoption of Big Data technologies	• **Lack of Big Data technologies** to test the value of new ways of processing big data and showcase its benefits to their management • **Labour-intensive manual approach** to extract information from huge volume of data (e.g. text mining) • **Lack of Big Data skills**	• **Facilitate the prototyping and launch of pilot projects on big data**, data analytics or text mining, by providing the infrastructure, the software tools and the associated support needed to start a pilot project • **Foster the use of open source software** • **Facilitate training programs on benefits** of the big data analytics	**Start**
Sharing of Big Data sources and test datasets	• **Data sharing among PAs is not yet a common practice** and the shared open datasets are often just aggregated data or samples	• **Provide built-in connectors/APIs for the ingestion and analysis of datasets made** available by European PAs / Institutions and several testing datasets classified into different policy domains in order to give the opportunity to experiment and discover useful and interesting correlations among data • **Foster the sharing of various data sources across domains and organisations to support better policy-making**	**Experiment**
Sharing of Big Data artefacts	• **Risk of replicating the efforts by implementing similar use cases**, which does not profit from the opportunities of sharing statistical models, algorithms, data connectors and methodologies already implemented • **Sharing of good practices, knowledge, experiences among different PAs** is not yet common practice	• **Support PAs through the creation of a community** around best practices and methodologies on big data for policy-making • **Offer a Big Data sandbox environment** with ready-to-use Big Data open analytics tools, statistical models, algorithms, data connectors and test datasets	**Share**

Agenda

Big Data Test Infrastructure

Background & Objectives

Approach and status of the study

Task 1 - Use cases and requirements analysis

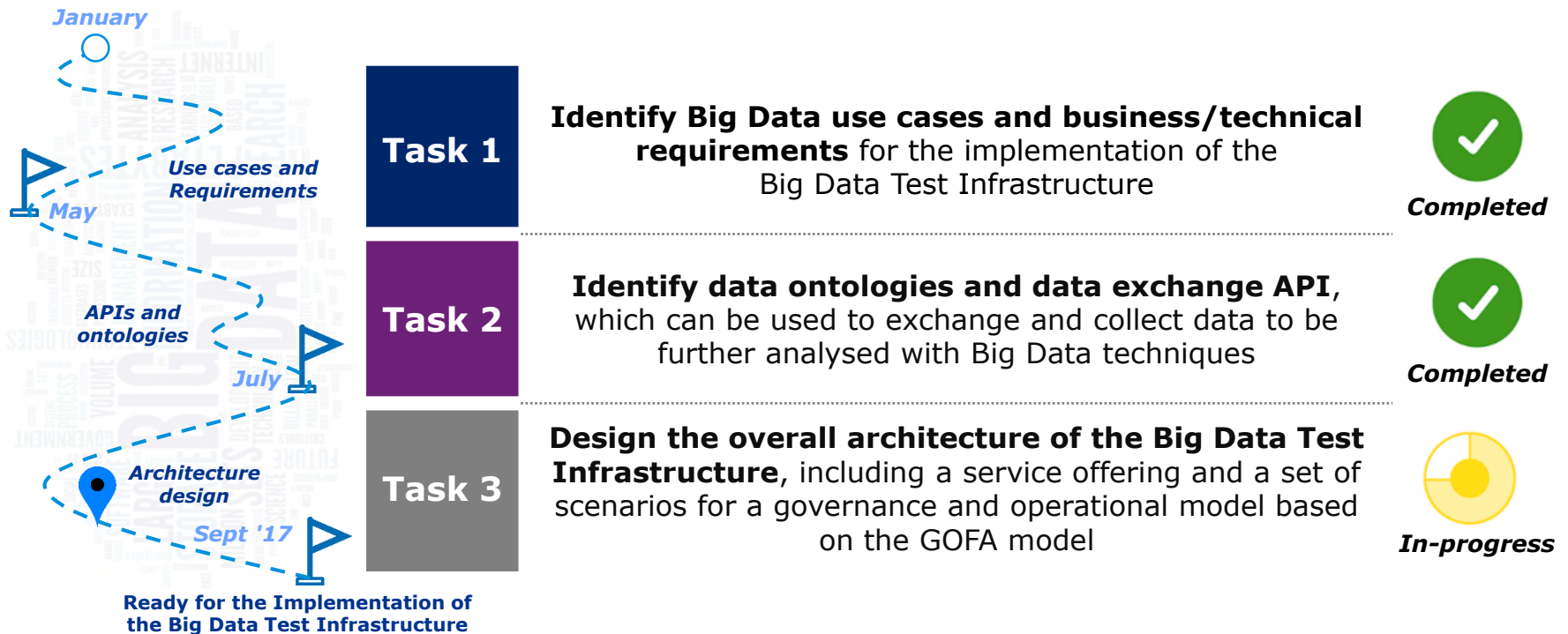Task 2 - Data ontologies and data exchange APIs

Task 3 - Design of the infrastructure

Next Steps

Big Data Test Infrastructure
Approach and status of the study

The Study "Big Data Test Infrastructure" is on-going and TASK 1 and Task 2 are finished.

January

Use cases and Requirements

May

APIs and ontologies

July

Architecture design

Sept '17

Ready for the Implementation of the Big Data Test Infrastructure

Task 1	**Identify Big Data use cases and business/technical requirements** for the implementation of the Big Data Test Infrastructure	✅ **Completed**
Task 2	**Identify data ontologies and data exchange API**, which can be used to exchange and collect data to be further analysed with Big Data techniques	✅ **Completed**
Task 3	**Design the overall architecture of the Big Data Test Infrastructure**, including a service offering and a set of scenarios for a governance and operational model based on the GOFA model	**In-progress**

Strong participation of Member States through the intermediary of the ISA Coordination Group. MSs on-board so far are: Slovenia, Spain, Portugal, Norway, Estonia, Malta, Czech Republic, Hungary and Netherlands

European Commission

Agenda

Big Data Test Infrastructure
Task 1 - Use cases and requirements analysis

Task 1 of the study has been completed in June '17 and the following action has been carried out:

- **Identification of Big Data use cases** to be supported by the future Big Data Test Infrastructure.

- **Identification of a set of business and technical requirements** that will guide the design of the future Big Data Test Infrastructure.

- **Identification of good practices coming from existing Big Data infrastructures** set up by public administrations at national level.
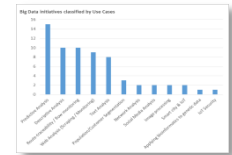
Data collection Activities	Outcomes
Bilateral Conferences with Member States	• **Emerging business needs** on the Big Data field from the interested Member States
Targeted Interviews with key stakeholders and Desk Research	• Technical analysis of existing **Big Data pilots (which use a Big Data infrastructure)** performed at EU and National level • **Additional needs** coming from relevant Big Data pilots
Final workshop with Member States	• **Refinement of the business and technical requirements** • **Consolidation of the entire study** thanks to several MSs' feedbacks

European Commission

Big Data Test Infrastructure
Task 1 - Use cases and requirements analysis

BIG DATA USE CASES
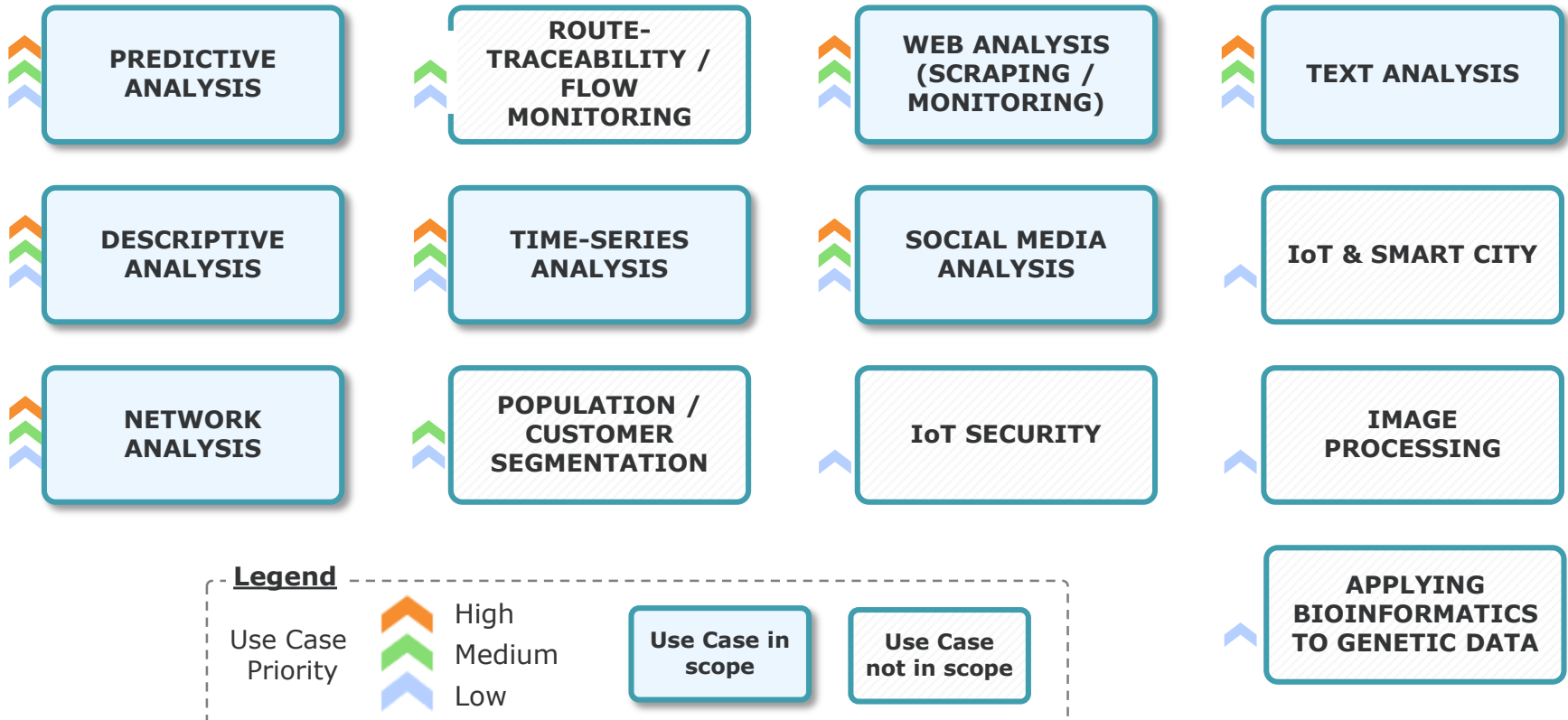
| PREDICTIVE ANALYSIS | ROUTE-TRACEABILITY / FLOW MONITORING | WEB ANALYSIS (SCRAPING / MONITORING) | TEXT ANALYSIS |

| DESCRIPTIVE ANALYSIS | TIME-SERIES ANALYSIS | SOCIAL MEDIA ANALYSIS | IoT & SMART CITY |

| NETWORK ANALYSIS | POPULATION / CUSTOMER SEGMENTATION | IoT SECURITY | IMAGE PROCESSING |

APPLYING BIOINFORMATICS TO GENETIC DATA

Legend

Use Case Priority
- High
- Medium
- Low

Use Case in scope

Use Case not in scope

European Commission

Big Data Test Infrastructure
Task 1 - Use cases and requirements analysis

During Task 1, the team collected a **set of business and technical requirements** structured in the following "areas" of the Technical framework adopted:

AREAS:
- **Governance & Security**: includes whatever is related to the provisioning, the management, the monitoring of the infrastructure and also the privacy and the security of infrastructure and data

- **Data Ingestion / Storage**: includes whatever is related to the acquisition of data and how it is stored and organised in the infrastructure

- **Data Elaboration**: includes whatever is related to the data processing, for example from simple ETL to advanced machine learning algorithms

- **Data Consumption**: includes whatever is related to the data expending for example in order to visualise it, create dashboard, export data

- **Infrastructure**: covers the aspects related to the technology stacks, including storage, servers computing and networking capacity
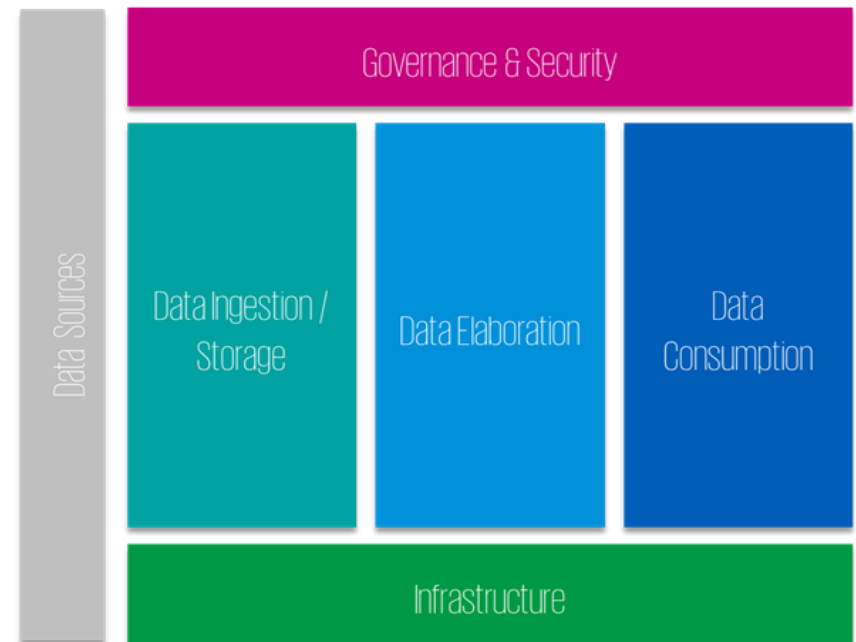
Framework used to classify the technical requirements

European Commission

Big Data Test Infrastructure
Task 1 - Use cases and requirements analysis

During Task 1, all the **recommendations and guidelines** identified from the analysis of relevant Big Data Infrastructures at EU and National level, to support the design of the Big Data Test Infrastructure.

For each **good practice**, the following **three dimensions of the sGOFA model** (CEF) have been taken into account:

- **Architecture dimension**
- **Operation dimension**
- **Governance & Skills dimension**

Good practice	Organisation
JRC Earth Observation & Social Sensing Big Data Pilot Project	JRC, EU level
Procurement of a Big Data platform through a framework contract for Italian PAs	CONSIP / AgID, National level
SANDBOX	UNECE, EU level
Big Data Cooperational System	ANAC, National level
Eurostat big data cluster	ESTAT, EU level
Amsterdam innovation ArenA – Data Driven Operator	City of Amsterdam, National level
BigData@Polito Cluster	Polytechnic of Turin, National level

GOOD PRACTICE EXAMPLE

European Commission

Agenda

Big Data Test Infrastructure
Data ontologies and data exchange APIs

Task 2 of the study has been completed in July '17. The team analysed the following potential data sources of the Big Data Test Infrastructure:

European Data Portal
EUROPEAN DATA PORTAL

API Rest, SPARQL-based queries, Geospatial query protocols, Web Services

EU Open Data Portal

API Rest, SPARQL-based queries, Geospatial query protocols, Web Services

National and international open APIs

Type of APIs depending on the specific data source

Social Media
twitter
facebook

Twitetr: Rest and Streaming API
Facebook: Graph API and Public Feed API

Data Sources

Governance & Security

Data Ingestion / Storage

Data Elaboration

Data Consumption

Infrastructure

The data sources are also used to individualize a set of **ontologies** and **datasets with associated APIs, as well as specific evaluation criteria** which will facilitate the selection of datasets for the implementation of Big Data pilots by Public Administration and Institutions at National and European level

Agenda

Big Data Test Infrastructure
Task 3 - Design of the infrastructure

The design of the infrastructure is currently undergoing. The methodology followed for the design is shown in the following figure:

Step 3.1.
Identification of the User Stories

Step 3.2.
Design of the target architecture

Step 3.3.
Design of the Governance and Operational model

Step 3.4.
Report on final results and Webinar

Identification of:

- **Users personas** (target personas for the usage of the Infrastructure)
- **Epics** (high-level use cases for the usage of the Infrastructure)
- **User stories**

Identification of:

- **Architecture principles** (ref. European Interoperability Framework)
- **Business services** the Big Data Test infrastructure will provide
- **Target architecture** (TOGAF style) with functional and solution Building blocks

Identification of **governance and operational scenarios** of the Big Data Test Infrastructure, which involves the aspects of control and management of the platform.

A **final webinar** will be organised with all involved stakeholders once the design will be completed

Estimated end of the study: end of September '17

European Commission

Agenda

Big Data Test Infrastructure

Background & Objectives

Approach and status of the study

Task 1 - Use cases and requirements

Task 2 - Data ontologies and data exchange APIs

Task 3 - Design of the infrastructure

Next Steps

Big Data Test Infrastructure
Next Steps

[End of September]

Sharing of the final results of the study with Member States

During the bilateral conferences with MSs, we have collected several business needs and potential requirements. Then we have presented in June the final result of Task 1 which received fruitful feedbacks. In the next meeting will be discussed:

- **Presentation of the results of Task 2 and 3**
- **Collection of MSs' feedbacks**

[Begin of 2018]

Implementation of the Big Data test Infrastructure

A small scale implementation has been scheduled for the begin of 2018 with the possibility to scale-up the entire project in the 2019.

Furthermore, Member States have expressed significant interest for the CEF candidate Building Block "Big Data Test Infrastructure" and there is an high probability to include the initiative in the 2018 CEF work programme as a fully mature Building Block.

Funding Idea for the future implementation of the Platform

- **2018: small scale** *isa* + CEF
- **2019: large scale** CEF

18

Thank you for your attention

BIG DATA TEST INFRASTRUCTURE

European Commission

DG INFORMATICS (DIGIT)

Contact us

Marco.FICHERA@ec.europa.eu